

General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

MATRIX DIFFERENTIATION FORMULAS

D. A. Usikov and D. K. Tkhabisimov



Translation of "Formuly Matrichnogo Differentsirovaniya",
Academy of Sciences USSR, Institute of Space Research,
Moscow, Report Pr-755, 1982, pp. 1-11.

(NASA-TM-77210) MATRIX DIFFERENTIATION
FORMULAS (National Aeronautics and Space
Administration) 13 p HC A02/MF A01 CSCL 12A

N83-3 1391

Unclas
G3/64 28399

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON, D.C. 20546 APRIL 1983

STANDARD TITLE PAGE

1. Report No. NASA TM-77210	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle MATRIX DIFFERENTIATION FORMULAS		5. Report Date April 1983	
		6. Performing Organization Code	
7. Author(s) D. A. Usikov and D. K. Tkhabisimov		8. Performing Organization Report No.	
		10. Work Unit No.	
9. Performing Organization Name and Address SCITRAN Box 5456 Santa Barbara, CA 93108		11. Contract or Grant No. NAS- 3542	
		12. Type of Report and Period Covered Translation	
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D.C. 20546		14. Sponsoring Agency Code	
13. Supplementary Notes Translation of "Formuly Matrichnogo Differentsirovaniya", Academy of Sciences USSR, Institute of Space Research, Moscow, Report Pr-755, 1982, pp. 1-11.			
14. Abstract A compact differentiation technique (without using indexes) is developed for scalar functions that depend on complex matrix arguments which are combined by operations of complex conjugation, transposition, addition, multiplication, matrix inversion and taking the direct product. The differentiation apparatus is developed in order to simplify the solution of extremum problems of scalar functions of matrix arguments. ORIGINAL PAGE IS OF POOR QUALITY			
17. Key Words (Selected by Author(s))		18. Distribution Statement Unclassified - Unlimited	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 13	22. Price

Matrix Differentiation Formulas

D. A. Usikov, D. K. Tkhabisimov
Space Research Institute, USSR Academy of Sciences

Submitted for Publication by
Deputy Director V. M. Balebanov

A compact differentiation technique (without using indexes) is developed for scalar functions that depend on complex matrix arguments which are combined by operations of complex conjugation, transposition, addition, multiplication, matrix inversion and taking the direct product. The differentiation apparatus is developed in order to simplify the solution of extremum problems of scalar functions of matrix arguments.

Often in solving the problems of mathematical statistics, the theory of probability, system theory and others it becomes necessary to calculate the derivatives of functions that depend on matrix arguments. If the relation to the arguments is expressed by a complex function, the calculation of the derivatives by development of the matrix operations in indexes becomes extremely laborious. In [1] a number of "recipes" are proposed that heighten the "probability" of obtaining a finite formula in the matrix expression. One of the authors of the present work has previously pointed out a general method of solving the problem without recourse to indexes [2]. The present work solves this problem in its entirety.

/3*

1. The General Relations of Matrix Differentiation

The basis of preserving the matrix form of the operation of differentiation with respect to a matrix argument is the

* Numbers in the margin indicate pagination in the foreign text.

special form of notation of the derivatives. Let $f(x)$ be a scalar function of the matrix argument x ; x is a matrix of size $m \times n$. We shall introduce the following designation:

$$\frac{\partial f}{\partial x} \equiv \begin{vmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{21}} & \cdots & \frac{\partial f}{\partial x_{m1}} \\ \frac{\partial f}{\partial x_{12}} & \frac{\partial f}{\partial x_{22}} & \cdots & \frac{\partial f}{\partial x_{m2}} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f}{\partial x_{1n}} & \frac{\partial f}{\partial x_{2n}} & \cdots & \frac{\partial f}{\partial x_{mn}} \end{vmatrix} \quad (1)$$

We shall define the operator form of notation for the differentiation of complex functions as follows. Let $y(x)$ be a matrix that depends on the matrix x , $f(y)$ be a scalar function of the matrix argument. By the rule of differentiation of complex functions we have:

$$\frac{\partial f(y(x))}{\partial x_{me}} = \sum_{ip} \frac{\partial f(y)}{\partial y_{ip}} \frac{\partial y_{ip}}{\partial x_{me}} \quad (2)$$

We shall rewrite relation (2) in the operator form:

$$\frac{\partial f}{\partial x} = \left[\frac{\partial y}{\partial x} \right] \frac{\partial f}{\partial y}, \quad (3)$$

where $\partial f / \partial x$ and $\partial f / \partial y$ are the matrix forms of notation of the partial derivatives as defined in definition (1); the expression in brackets here and in future shall denote an operator which by definition operates on the matrix to the right (in the present case, $\partial f / \partial y$). If the complex function has the form $f(z)$,

where $z = z(y)$, $y = y(x)$, z, y, x being matrices, then analogous with (2) we have:

$$\frac{\partial f(z)}{\partial x_{me}} = \sum_{kn} \frac{\partial y_{kn}}{\partial x_{me}} \sum_{ip} \frac{\partial z_{ip}}{\partial y_{kn}} \frac{\partial f(z)}{\partial z_{ip}}. \quad (4)$$

or, writing (4) analogous to (3) we obtain in the operator form:

$$\frac{\partial f}{\partial x} = \left[\frac{\partial y}{\partial x} \right] \left[\frac{\partial z}{\partial y} \right] \frac{\partial f}{\partial z} \quad (5)$$

In the general case, if the complex function has the form:

$$f(z_1(\dots(z_q)))$$

then the following formula is obviously valid:

$$\frac{\partial f}{\partial z_q} = \left[\frac{\partial z_{q-1}}{\partial z_q} \right] \dots \left[\frac{\partial z_1}{\partial z_2} \right] \frac{\partial f}{\partial z_1} \quad (R0)$$

Formula (R0) reduces the initial problem to one of operation by operators:

$$\left[\frac{\partial z_{q-1}}{\partial z_q} \right]$$

We primarily observe the obvious relation:

$$\frac{\partial f}{\partial x} = \left[\frac{\partial x}{\partial x} \right] \frac{\partial f}{\partial x} \quad (R1)$$

The following relations enable a notation of the derivative in matrix form if the argument of function f is obtained

from matrices by the operations of complex conjugation, transposition, matrix inversion (hereafter we write $F(y) = \partial f(y)/\partial y$):

$$\frac{\partial f(\tilde{x})}{\partial x} = \left[\frac{\partial(\tilde{x})}{\partial x} \right] F(y) = \tilde{F}(y) \Big|_{y=\tilde{x}}, \quad (R2)$$

$$\frac{\partial f(x^T)}{\partial x} = \left[\frac{\partial(x^T)}{\partial x} \right] F(y) = (F(y))^T \Big|_{y=x^T}, \quad (R3)$$

$$\frac{\partial f(x^{-1})}{\partial x} = \left[\frac{\partial(x^{-1})}{\partial x} \right] F(y) = -y^* F(y) y \Big|_{y=x^{-1}}, \quad (R4)$$

The proofs of relations (R2)-(R4) are done in indexes. As an example we carry out the proof of (R4). By definition of an inverse matrix:

$$\sum_j (x^{-1})_{ij} x_{jk} = \delta_{ik} \quad (6)$$

which is the Kronecker symbol. We differentiate (6) with respect to x_{me} :

$$\frac{\partial}{\partial x_{me}} \sum_j (x^{-1})_{ij} x_{jk} = \sum_j \frac{\partial (x^{-1})_{ij}}{\partial x_{me}} x_{jk} + \sum_j (x^{-1})_{ij} \delta_{jm} \delta_{ke} = 0 \quad (7)$$

In the second term of (7) we can sum over j . Then we multiply (7) by:

$$(x^{-1})_{kp}$$

and sum over all k :

$$\sum_{ik} \frac{\partial (\bar{x}^{-1})_{ij}}{\partial x_{me}} x_{jk} (\bar{x}^{-1})_{kp} + \sum_k (\bar{x}^{-1})_{im} (\bar{x}^{-1})_{kp} \delta_{ke} = 0 \quad (8)$$

Taking account of (6) and summing over k in the second term, we obtain from (8):

$$\frac{\partial (\bar{x}^{-1})_{ip}}{\partial x_{me}} = -(\bar{x}^{-1})_{im} (\bar{x}^{-1})_{ep}$$

We shall now prove (R4) in the index form:

$$\frac{\partial f(\bar{x}^{-1})}{\partial x_{me}} = \sum_{ip} \frac{\partial f(\bar{x}^{-1})}{\partial (\bar{x}^{-1})_{ip}} \frac{\partial (\bar{x}^{-1})_{ip}}{\partial x_{me}} = - \sum_{ip} (\bar{x}^{-1})_{im} \frac{\partial f(y)}{\partial y_{ip}} (\bar{x}^{-1})_{ep} \quad (9)$$

Taking note of the form of notation (1), we obtain (R4) in an obvious manner from (9) in the matrix form. The proof is complete. /6

The next two relations, easily proved in indexes, allow us to reduce the calculation of the derivatives of complex functions to more simple expressions and, together with relations (R0)-(R4), allow a calculation in matrix form of the derivatives of any given matrix expressions. Let A(x) and B(x) be matrices that are functions of the matrix argument x; then:

$$\frac{\partial f(AB)}{\partial x} = \left[\frac{\partial AB}{\partial x} \right] F(y) = \left[\frac{\partial A}{\partial x} \right] B F(y) + \left[\frac{\partial B}{\partial x} \right] F(y) A \Big|_{y=AB}; \quad (R5)$$

$$\frac{\partial f(A+B)}{\partial x} = \left[\frac{\partial (A+B)}{\partial x} \right] F(y) = \left[\frac{\partial A}{\partial x} \right] F(y) + \left[\frac{\partial B}{\partial x} \right] F(y) \Big|_{y=A+B} \quad (R6)$$

As an example of the use of formulas (R0)-(R6) we shall calculate the derivative:

$$\frac{\partial f}{\partial x} f(Q(x^T A x)^{-1}).$$

We denote: $y = x^T$,

$$z = Ax, p = yz, q = p^{-1}, r = Qq.$$

Step 1. Per (R0): $\frac{\partial f}{\partial x} = \left[\frac{\partial r}{\partial x} \right] F.$

Step 2. Per (R5): $\left[\frac{\partial r}{\partial x} \right] F = \left[\frac{\partial(Qq)}{\partial x} \right] F = \left[\frac{\partial Q}{\partial x} \right] q F + \left[\frac{\partial q}{\partial x} \right] F Q = \left[\frac{\partial q}{\partial x} \right] F Q.$

Step 3. Per (R4): $\left[\frac{\partial q}{\partial x} \right] F Q = \left[\frac{\partial p^{-1}}{\partial x} \right] F Q = \left[\frac{\partial p}{\partial x} \right] \left[\frac{\partial p^{-1}}{\partial p} \right] F Q = \left[\frac{\partial p}{\partial x} \right] m.$

where for brevity in writing the further calculations we denote:

$$m = -p^{-1} F Q p^{-1}.$$

Step 4. Per (R5): $\left[\frac{\partial p}{\partial x} \right] m = \left[\frac{\partial(yz)}{\partial x} \right] m = \left[\frac{\partial y}{\partial x} \right] z m + \left[\frac{\partial z}{\partial x} \right] m y.$

Step 5. Per (R3): $\left[\frac{\partial y}{\partial x} \right] z m = \left[\frac{\partial x^T}{\partial x} \right] z m = (z m)^T = m^T z^T = m^T x^T A^T.$

Step 6. Per (R5): $\left[\frac{\partial z}{\partial x} \right] m y = \left[\frac{\partial Ax}{\partial x} \right] m y = \left[\frac{\partial A}{\partial x} \right] x m y + \left[\frac{\partial x}{\partial x} \right] m y A = m x^T A.$

Finally, we get:

$$\begin{aligned} \frac{\partial f}{\partial x} &= m^T x^T A^T + m x^T A = -(p^{-1} F Q p^{-1})^T x^T A^T - p^{-1} F Q p^{-1} x^T A = \\ &= -(x^T A x)^{-1} Q^T F^T (x^T A x)^T x^T A^T - (x^T A x)^{-1} F Q (x^T A x)^{-1} x^T A. \end{aligned} \quad (10)$$

If the reader still has doubts about the usefulness of the matrix method of differentiation, it is worthwhile to try to obtain formula (10) by calculations with indexes. /7

We shall fill out the list of permissible matrix operations with the direct product of matrices and derive the corresponding differentiation formulas. The direct product of A and B in the notation:

$$A \otimes B$$

is defined as the block matrix:

$$A \otimes B = \begin{vmatrix} A_{11}B & A_{12}B & \dots & A_{1m}B \\ A_{21}B & A_{22}B & \dots & A_{2m}B \\ \dots & \dots & \dots & \dots \\ A_{k1}B & A_{k2}B & \dots & A_{km}B \end{vmatrix}. \quad (11)$$

If matrices A and B are of respective size $k \times m$ and $r \times s$, according to definition (11):

$$(A \otimes B)_{(i-1)r+p, (j-1)s+e} = A_{ij} B_{pe}.$$

Let A and B be matrices that depend on the matrix x, $f(A \otimes B)$ be a scalar function of the matrix $A \otimes B$; by definition: /8

$$F(y) = \frac{\partial f(y)}{\partial y} \Big|_{y=A \otimes B}$$

We obtain an analog of formula (R5) for the case of the direct product:

$$\begin{aligned} \frac{\partial f(A \otimes B)}{\partial x_{nq}} &= \sum_{ijpe} F_{(j-1)s+l, (i-1)r+p} \frac{\partial A_{ij} B_{pe}}{\partial x_{nq}} = \\ &= \sum_{ijpe} F_{(j-1)s+l, (i-1)r+p} \left(\frac{\partial A_{ij}}{\partial x_{nq}} B_{pe} + A_{ij} \frac{\partial B_{pe}}{\partial x_{nq}} \right) = \quad (12) \\ &= \sum_{ij} \frac{\partial A_{ij}}{\partial x_{nq}} \sum_{pe} B_{pe} F_{(j-1)s+l, (i-1)r+p} + \sum_{pe} \frac{\partial B_{pe}}{\partial x_{nq}} \sum_{ij} F_{(j-1)s+l, (i-1)r+p} A_{ij}. \end{aligned}$$

Designating in (12):

$$(B \odot F)_{ij} = \sum_{pe} B_{pe} F_{(j-1)s+l, (i-1)r+p}, \quad (13)$$

$$(F \odot A)_{pe} = \sum_{ij} F_{(j-1)s+l, (i-1)r+p} A_{ij}. \quad (14)$$

By its meaning, the operation \odot is an operation that is inverse to the taking of the direct product. The following formula obtains in designations (13) and (14):

$$\frac{\partial f(A \otimes B)}{\partial x} = \left[\frac{\partial A}{\partial x} \right] B \odot F(y) + \left[\frac{\partial B}{\partial x} \right] F(y) \odot A. \quad (R7)$$

2. Differentiation of the Trace and the Determinant

The most common functions of a matrix argument are the type $\text{tr}(x)$ and $\det(x)$ (defined only for the square matrices x),

which it is worth giving a special consideration. In accordance with definition (1):

$$\frac{\partial \text{tr}(x)}{\partial x} = I \quad (\text{T1})$$

where I is the unit matrix. In particular, in formula (10) if f is the trace, $F = I$.

9

Using relations (R4) and (R5) we obtain the formula:

$$\frac{\partial \text{tr}(x^n)}{\partial x} = n x^{n-1} \quad (\text{T2})$$

which is valid for $n = 0, \pm 1, \pm 2, \dots$

Let $F(x)$ be a matrix function of a matrix argument, which can be represented in the form of the Taylor series:

$$F(x) = \sum_{k=0}^{\infty} f_k x^k$$

Then, using (T2), we determine the matrix derivative:

$$\frac{\partial \text{tr} F(x)}{\partial x} = \frac{\partial}{\partial x} \sum_{k=0}^{\infty} f_k \text{tr}(x^k) = \sum_{k=0}^{\infty} f_k k x^{k-1} \quad (\text{T3})$$

In particular:

$$\frac{\partial \text{tr}(e^x)}{\partial x} = e^x \quad (\text{T4})$$

To obtain the formulas for the derivatives of a determinant we observe that the determinant can be represented as the sum of any given row [3]:

$$\det(x) = \sum_j x_{ij} X_{ij}$$

where:

$$X_{ij}$$

is the algebraic complement of the element x_{ij} and, consequently,

X_{ij} does not depend on x_{ij} . Thus:

$$\frac{\partial}{\partial x} \det(x) = \begin{vmatrix} X_{11} & X_{21} & \dots & X_{n1} \\ X_{12} & X_{22} & \dots & X_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1n} & X_{2n} & \dots & X_{nn} \end{vmatrix}.$$

If $\det(x) \neq 0$, it is known that a matrix comprised of algebraic complements is expressed in terms of the inverse matrix [3], and therefore when $\det(x) \neq 0$:

$$\frac{\partial \det(x)}{\partial x} = x^{-1} \det(x). \quad (D1)$$

We shall compute the derivative of $\det(x^n)$:

$$\frac{\partial \det(x^n)}{\partial x} = \frac{\partial (\det(x))^n}{\partial x} = (n \det^{n-1}(x)) (\det(x))^{-1} = n \det(x)^{n-1}. \quad (D2)$$

We derive a formula analogous to (T3) for the case of the determinant:

$$\frac{\partial \det(F(x))}{\partial x} = \frac{\partial}{\partial x} \det\left(\sum_{k=0}^{\infty} f_k x^k\right) = \sum_{k=0}^{\infty} \left[\frac{\partial f_k x^k}{\partial x}\right] (F(x)^{-1} \det F(x))$$

$$= \sum_{k=0}^{\infty} f_k k x^{k-1} F(x)^{-1} \det F(x) = F(x)^{-1} \det F(x) \sum_{k=0}^{\infty} f_k k x^{k-1} \quad (D3)$$

In deriving (D3) we made use of relations (R6) and (D2) and the fact that matrices $F(x)$ and x commute. In particular:

/10

$$\frac{\partial \det(e^x)}{\partial x} = (e^x)^{-1} \det(e^x) \sum_{k=0}^{\infty} \frac{k x^{k-1}}{k!} =$$

$$= (e^x)^{-1} \det(e^x) \sum_{k=1}^{\infty} \frac{x^{k-1}}{(k-1)!} = (e^x)^{-1} \det(e^x) \sum_{k=0}^{\infty} \frac{x^k}{k!} = \quad (D4)$$

$$= (e^x)^{-1} \det(e^x) e^x = I \det(e^x).$$

References

1. Bard, Y., Nelineynoye otsenivaniye parametrov [Non-linear Evaluation of Parameters], Statistika, Moscow, 1979.
2. Usikov, D. A., Invarianty podobiya kovariatsionnoy matritsy i kriterii optimal'nosti [Affine Invariants of a Covariant Matrix and the Criteria of Optimality], Preprint FEI-490, Obninsk, 1974.
3. Lankaster, P., Teoriya matrits [The Theory of Matrices], Nauka, Moscow, 1978.